DOCUMENT RESUME

ED 139 388 IR 004 666

AUTHOR Hitchingham, Eileen E.
TITLE On-Line Data Bases. II. Search Performance Evaluation and On-Line Data Bases.
SPONS AGENCY Council on Library Resources, Inc., Washington, D.C.
PUB DATE 75
NOTE 30p.

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Data Bases; Evaluation Methods; Information Retrieval; Literature Reviews; *Measurement Techniques; *On Line Systems; *Performance Criteria; *Relevance (Information Retrieval); *Search Strategies

ABSTRACT

In an attempt to measure on-line searching performance, test searches of one question conducted by three bibliographic searchers are analysed to determine the effects of search strategy on results. The study reviews retrieved citations for relevance, calculating precision values, and estimated recall for each search. Performance analysis reveals that the "best" search strategy can be determined using different criteria--low cost, better recall, or high recall. The report summarizes findings of previous studies measuring: (1) cost, terminal time, and user satisfaction; (2) precision; and (3) recall. It suggests the performance measure of unit cost per relevant citation determined by cost of terminal time plus cost of off-line prints divided by the number of relevant citations retrieved. Suggested formulas for determining recall values in high recall need searches are defined for institutions with one searcher and institutions with several searchers. (KP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED139388

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

On-Line Data Bases. II. Search Performance Evaluation and On-Line Data Bases.

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Eileen E. Hitchingham

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

by

Eileen E. Hitchingham
Oakland University
Rochester, Michigan 48063

A report to the Council
on Library Resources of
the 1974/75 Fellowship project

I R0047666

FII

Advances in on-line technology and the appearance of commercial vendors of on-line data bases have created a climate in which libraries can consider a new service option -- provision of computerized bibliographies tailored to the need of the individual requester. Cuadra (1) has indicated that 140 institutions were using SDC's ORBIT system in February 1974. In the academic sector I am aware of 43 institutions which provided, or anticipated providing on-line searches as of Summer 1974 (2). It is likely that a large number of institutions will be revealed as on-line users in the forthcoming report of an NSF sponsored study of on-line system use (3). Several reports of experiences with on-line systems have recently appeared in the literature (4,5,6, 7,8).

Some questions concerning the provision of on-line bibliographies by libraries appear worth examining. I consider the provision of these bibliographies to be a reference/information service, wherever the searching function is administratively housed in an institution. There are indications of considerable interest in measurement techniques that can be applied to traditional reference services (9). Is there a corresponding interest in measurement of this particular reference service -- provision of bibliographies by on-line searching? What measures are being applied to this new service?

Saracevic (10) considers the "secret" of performance for on-line systems to be skillful question negotiation i.e. the method of analysis applied to the question and the construction of logical statements for searching. The effects of various search strategies on the results for a particular question are illustrated in the following test search. The discussion section considers performance measures that have been applied to non-experimental on-line services. It suggests the need for additional measures -- measures

recognizing the effectiveness and efficiency of search strategies.

TEST SEARCH

A faculty member in the Chemistry Department of Oakland University prepared a short narrative statement for a search request to be run against the Chemical Abstracts (CA) Condensates data base. In the request he indicated interest in "specific chemical compounds found in insects." He explained further in the letter that titles of articles relevant to his request might refer to a specific chemical compound or a class of chemical compounds "e.g., composition of the faas of spiders." He indicated that he could be contacted by phone for additional information.

The data base includes all of the citations which are available in the printed version of Chemical Abstracts. CA Condensates is available through both Lockheed and SDC and is searched by batch process at a number of tape-leasing information centers. The main approach to the subject is through uncontrolled keyword indexing.

Two of the searchers participating in the test were located in academic libraries recently involved in on-line searching; the third searcher was employed by a search center established in the late 1960's. The center is primarily organized to provide current awareness searches, but does offer a retrospective search service.

Searchers at the academic institutions had agreed to participate in the test when I interviewed them as part of the previously mentioned study of on-line services in academic libraries. A third searcher who had expressed interest when interviewed, declined to participate after receiving the mailed request,characterizing the request as "too general to be a fair test. The entry as given in ChemCon is specific. Answering the question would

entail entering every single insect name. The same is true of the chemical compounds, unless one assumed that anything concerning insects listed in ChemCon would also concern chemical compounds." This very neatly summarizes the difficulties inherent in this particular question and similar questions which involve determining synonymous terms to express a particular concept of a request in a free-text and/or non-hierarchical system (e.g. the difficulty of expressing the concept "psychological effects" in a request for a bibliography on "psychological effects of institutionalization on the aged"). At the same time the question is real, serving a real information need -- background for a review publication. The same difficulties would be present and compounded in a manual search of Chemical Abstracts.

The search from the center was limited to volumes 78 and 79 (1973) of Chemical Abstracts. At the time the request was submitted volumes 80 and 81 were being searched in a current awareness mode only and would not be available at retrospective search rates until 1975. The on-line searches covered the CA Condensates data base from 1972 through Fall 1974 (volumes 76-81).

Searcher A had seven months experience with CA Condensates and had conducted four searches on the data base during that time; Searcher B had 11 months experience and had done 70 searches with the Condensates; searchers at the center have been characterized by the director as "very well experienced."

Search Performance Characteristics

Searches A and B were received within 10 days after the two individuals agreed to participate. With Search C the faculty member was contacted three months after his initial request. Search results were received one month later. The center was approached as a commercial establishment and was

therefore given no indication that this was a test search. It is suspected that turn-around time was an exception to the norm since faster response from the center has been reported elsewhere (11).

Search A retrieved 544 unique citations and was invoiced at $120.32; Search B retrieved 247 unique citations and was invoiced at $30.21; Search C retrieved 1324 citations and was invoiced for $145.00. Components of the charges might be considered. Search A results from two terminal passes on two consecutive days. On the first day 42 minutes were spent at the terminal to retrieve postings for 636 citations of which 25 were printed off-line. A call to clarify whether all citations would be acceptable (concern for the cost of 600+ citations) and a positive answer to this question led to the input of essentially the same search the next day in 26 minutes and the retrieval of postings for 544 citations which were printed off-line. Since an error in the Boolean logic was noted and corrected during the second input, total time for the search is considered to be 68 minutes. Institution A charged $1.10 per minute of terminal time and eight cents per citation for off-line prints.

Search B required 11 minutes of terminal time to retrieve and print off-line 247 unique citations. Institution B charges $.95 per minute of terminal time and eight cents per citation for off-line prints. Search C was a batch search of two CA Condensates volumes (78 - 79) at $80.00 for the two volumes searched plus $65.00 for the 1324 citations sent. Table 1 (page 23) details characteristics and costs for the searches.

Search Results

Citations from the three searches were reviewed by the requester for relevance. Precision values for each search were calculated according to

the formula:

$$\text{Precision} = \frac{\text{Number of relevant citations retrieved}}{\text{Total number of citations retrieved}}$$

Search A had a value of 18.6% (101 relevant in 544 retrieved); Search B had a precision value of 12.6% (31 relevant in 247 retrieved); Search C had a value of 5.5% (73 relevant in 1324 citations retrieved).

An estimated recall base was established as the sum of the unique relevant citations from all searches (151 citations) plus 17 additional references which were part of a file the faculty member had maintained on the subject over the past several years. A citation for each of the 17 references could be located in the printed volumes of Chemical Abstracts corresponding to the volume coverage of the data base (vols. 76 - 81) so it was assumed that they should have been available on CA Condensates. The overall recall base was thus considered to consist of 168 citations.

Estimated recall values for each search were determined according to the formula:

$$\text{Estimated recall} = \frac{\text{Number of relevant citations retrieved in the search}}{\text{Total number of unique relevant citations from the three searches + unique citations known to the requester}}$$

Search A retrieved 101 relevant citations out of a possible 168 for a recall value of 60.1%; Search B retrieved 31 relevant citations out of 168 for a recall value of 18.5%; Search C retrieved 73 relevant citations out of 168 for a recall value of 43.5%.

The preceding values however provide a distorted comparison since Search C was limited to the two volumes for 1973. Comparison of results in searching these two volumes presents a somewhat different picture. Search A retrieved 23 relevant citations in 181 retrieved, for a precision value of

12.7%. Seventy-five citations were known to be relevant in these two volumes so adjusted recall for Search A for the one year is 30.7%. Search B retrieved 3 relevant references in 64 retrieved, giving an adjusted precision value of 4.7% and an adjusted recall value of 4%. Search C has the same precision value previously mentioned but retrieved 73 out of 75 known relevant citations. This gives Search C an adjusted recall value of 97.3%.

A unit cost in minutes of terminal time per relevant citation retrieved has been considered as a possible measure of efficiency of on-line systems (12,13). If this value is calculated for the two on-line searches a value of .67 minutes per relevant citation (68 minutes ÷ 101 relevant citations) is shown for Search A. Search B has a value of .36 minutes per relevant citation (11 minutes ÷ 31 relevant citations).

From the user's viewpoint a different measure of efficiency might be considered, actual cost (invoiced charges) per relevant citation in relation to recall. This value takes the cost of charges incurred for printing of irrelevant citations into account as well as terminal efficiency. It indicates cost per relevant citation at a recall level. Search A has a cost of $1.19 per relevant citation at 60.1% recall; Search B's cost is $.97 per relevant citation at 19.4% recall; Search C is $1.99 per relevant citation at 43.5% recall. Table 2 (page 24) gives overall precision and estimated recall values; adjusted precision and recall values for 1973 (vols. 78-79); unit costs in minutes of terminal time per relevant citation; and unit costs in invoiced charges per relevant citation.

Search Strategy Performance Analysis

Each of the searchers called the user to discuss the question. The formulation for Search A involves three concepts: Insects (I), Chemical

Composition (II), and Negated Elements (III). Search B's formulation involves two concepts: Insects (I), and Chemical Composition (II). Title searching for citations in the data base was an option available at the time the searches were conducted; neither Search A nor Search B employed this feature. Terminology and strategy for Search A is given in Figure 1 (page25), for Search B in Figure 2, (page 26).

The formulation for Search C is not available, but it appears to be of the type considered by the searcher who declined participation i.e. a search on all the terms that could be included for the concept Insects (I) with the supposition that everything in the Condensates data base would relate to chemical composition. Retrieval of such references as V78 #93414 entitled "Coffee Insects" with the keywords INSECT and REVIEW, and V78 #162187 entitled "First Find of Diabase in the PreCambrian of the Middle Bug River," with the keywords AMPHIBOLIZED and GABBRO fail to indicate a requirement for the intersection of an insect term with any term that might indicate chemical composition.

Search C achieved a 97.3% recall rate for volumes 78-79 at the expense of precision. It appears that Searcher C did not ascertain that the user was not interested in enzymes, DNA, RNA, or insecticides, concepts that could have been negated as in Search A. Negation of just those elements included in Search A would have removed 170 irrelevant citations (Negation of CA publication section 007 Enzymes; RNA, DNA, ENZYM#, INSECTICID#, PESTICID#). Expansion of negative concepts to include such things as ASE, a suffix in the named enzymes; CONTROL; Patents; CHEMOSTERIL#; ANALOG#; REPELLEN#; and STERIL#, and negating citations with the words "EFFECT" and "ON" would have further reduced irrelevant citations by 361 citations. The cumulative effect of such negations would have been the removal of 531 irrelevant citations

and the loss of 2 relevant citations. Although these negations would not have a great effect on the precision values (from 5.5% to 9%) the cost to the user would be reduced by approximately $27.00 or 18%.

Since the formulation for Search C does appear to have been very thorough in the listing of insect names the two unretrieved citations of the 75 known relevant in volumes 78-79 were examined (V78 #133669 and V78 #82216 from the requester's file). It appears that the first citation may have been missing from the tapes searched by the center since one of the keywords for the citation MOSQUITO does appear in a number of citations retrieved in Search C. The second citation may have been missed for the same reason, i.e. it was not on the tapes, or Search C did not include the term MILLIPEDE one of the keywords for this citation.

Search A has a somewhat better precision rate than Search C for volumes 78-79; however it also would have benefited by expansion of terms in the negated concepts. Including the additional terms suggested for Search C would have removed 59 irrelevant citations from volumes 78-79 and 150 irrelevant citations from the total retrieved. An even more interesting aspect is the consideration of terminology expansion which would have improved the recall performance for Search A in volumes 78-79. Searcher A has what might be considered the most complex formulation, a combination of 16 terms to express insects, and 30 items made up of word stems, abbreviations from the Chemical Abstracts list of abbreviated terms, and the numbers of publication sections of Chemical Abstracts to express chemical composition. Six items were used to indicate negated concepts. As previously indicated titie searching was an option in this search but was not used. Title searching in the system used required that a subset of the data base be created first;

a possible subset for this question would have been a request for section 012, NON MAMMALIAN BIOCHEMISTRY. The existing strategy with the inclusion of this feature would have retrieved only 2 of the 52 relevant citations not retrieved in Search A. Retrieval of 48 of the 52 missed references would have required searching of keywords and titles and the addition of 26 chemical and insect terms. Of the remaining four citations one is not in section 012 and could not be retrieved by a title search and 012 strategy, one was negated by ENZYM#, and two appear to not have been on the data base searched by A since they should have been retrieved by A's strategy. A recall level of 66 2/3% (50 out of 75 citations) could have been achieved by the addition of 9 terms (COMPOSITION, SECRETION, APHID#, SPIDER#, LEPIDOPTERA, LOCUST, LADYBUG, DROSOPHILA and OCCURRENCE) and title searching. Forty-seven additional irrelevant citations would have been retrieved by expansion to this strategy. It is interesting to note that while spiders are not true insects but rather are of the class Arachnida, inclusion of the terms suggested in the requester's example, SPIDER and COMPOSITION, and the use of title searching would have improved recall from the 30.7% achieved in the search to 41.3%. The searching of 16 Condensates publication sections as a way of expressing the chemical composition aspect retrieved one relevant citation in Search A that would not otherwise have been retrieved. This strategy was also responsible for the retrieval of 26 irrelevant citations.

It is recognized that even if the Searcher in A knew those terms which prove by hindsight to be the most productive, the additional input time for more terms and particularly the time involved in the title searching would have increased terminal time significantly. For example, title searching at that time with the data base used would have required: First, the creation of an insect term set and input of all terms that could have meant

chemical composition in the title (insect term in keyword and chemical composition term only in title). Second, creation of a chemical composition set and input of all terms that could have been used for insect in a title (chemical composition in keyword and insect term only in title). Third, creation of a subset of the Condensates data base, in this case 012, and input of all chemical composition terms and all insect terms that could appear in a title (insect term in title and chemical composition term in title). Fourth, a combination of insect set and chemical composition set already created in 1 and 2 (insect term in keyword, chemical composition in keyword). Negation of items would also require two inputs, once to negate in the keywords and a second time to negate in the title.

Search B retrieved 50 unique citations which were not retrieved in Search A or Search C. All of the unique citations were irrelevant. Eighteen were retrieved because of lack of negation (insecticid#, pesticid#, enzym#). Thirty-eight were retrieved by searching on protein#, amino and acid#, and antigen# -- terms unique to Search B's formulation.

Discussion

Which is the "best" search? is a difficult question. Each of the searches might be so considered dependent upon the criteria used to define "best." If low cost is the primary concern, Search B is "best" since it is the least costly overall and has the lowest unit cost per relevant citation. If a somewhat higher unit cost is acceptable to achieve better recall Search A is "best". It retrieved the greatest number of relevant citations, 3.3 times that of Search B and 1.4 times as many as Search C. Its unit cost per relevant citation is 16% greater than that of Search B. If high recall is the primary concern in searching, and a unit cost per relevant citation at

least double that of Search B is acceptable, Search C's strategy applied to three years of the on-line data base could be considered "best."

Information necessary for determining a "best" search in the sense of the preceding discussion, i.e. cost per relevant citation at levels of recall, has not been reported for on-line searching. Rather reports have concentrated on various elements of search performance, cost, terminal time, user satisfaction, precision, and recall, which only in conjunction with each other can provide a basis for search performance evaluation.

Cost, Terminal Time, User Satisfaction

Several reports dicuss gross average costs, terminal time used and user satisfaction. During a seven month period at EXXON average search costs were found to range from $12 to $38. Data bases used were in most cases scientific, and on-line time, off-line prints, and terminal rental are included in the cost. Elman (5) reports an average cost of $47 for on-line searching in a technical environment. The figure includes administrative costs as well as terminal time and print-out costs. An average cost of $6.44 has been reported for 56 on-line searches with ChemCon (14). In an academic environment MIT (7) reports total average cost figures comparable to those reported by Elman, $50.47 for data bases other than MEDLINE. Based on non-administrative costs, Hock (6) indicates average costs per search are less than $12. ERIC is the most popular data base. The Texas Information Service reports a similar average cost of $13 for ERIC searches performed on-line during a six-month period (15). This figure includes terminal rental, time on line, communication costs, and off-line prints.

At first glance information concerning costs would appear to be useful. One of the questions Cuadra (1) sees raised preliminary to contracting for

on-line services is "Can I afford it?" Reports of relatively low costs are reassuring to institutions with this question, but at the same time they invite dangerous comparisons at the management level. If an institution can provide searches at an average cost of $X, why is this institution providing searches for $X + Y ? The natural inclination to raise this question is shown in a study of on-line use of ERIC, conducted by the Institute of Library Research (16). In outlining the problem Bourne, et al. indicated:

> In studying the searching activity of the various ERIC service subscribers over a period of many months, it has been observed that wide variations exist among installations with respect to the average number of ERIC searches processed per unit of terminal time. Certain organizations consistently conduct more on-line searches per hour on their terminal than do others. Some installations consistently realize close to three times as many searches per hour as some other installations. This seems rather surprising considering that each installation searches the same data base, uses similar terminal equipment, receives similar instruction, uses similar support tools, and is served by the same central facility.

The primary objective of the study was to identify and explain the reasons for the differences in average on-line search rates between different terminal installations. In summarizing their study the investigators conclude their data indicate no single dominant factor that strongly influences the search speeds of the various installations -- rather that there is a complex pattern of influences. They feel that several factors seem to have influence on search speed -- work habits and formulation styles of searchers, presearch preparation for each question, cost-conscious attitude of searchers, searcher familiarity with the data base, and characteristics and speed of the terminal equipment. Only the influence of the last of these factors, terminal equipment, is supported by solid evidence from the study.

If the evaluation of on-line searching is to be more than a simple comparison of average costs for a particular data base, and if comparisons are not to be detrimental to higher costing searches (more terminal time per search),it appears that the problem of quality of the searches must be addressed. Few reports have touched upon evaluation of the quality of on-line searches other than to indicate that the user is satisfied or appears to be satisfied. User satisfaction is not a sufficient measure of search performance. As Carmon (17) suggests such satisfaction may be false or unwarranted in those cases where all retrieved answers are relevant but only a fraction of the relevant answers are retrieved. Rodgers (18) indicates that in some instances user satisfaction may be more a function of client naivete than of search quality. It seems then that evaluation of searching performance in providing on-line bibliographic services must include not only economic measures and an assessment of user satisfaction, but must also include those traditional measures which have previously been applied to experimental and operational batch information retrieval services, the measures of precision and recall.

Precision

An attempt to assess precision values, or the measure of the number of relevant citations retrieved by a search in relation to the total number of citations retrieved by the search, has been made in a few reports of experiences with on-line systems. A very rough measure of precision or search relevancy is indicated in the 36 questionnaires returned for 186 searches conducted by the Ontario Institute for Studies in Education during an experimental provision of on-line ERIC searches (19). Asked whether the printout

included particularly relevant citations, 2 respondents indicated NONE, 14 indicated A FEW, 10 indicated SEVERAL, and 10 indicated MANY. Estimated average cost for the searches was $26.80. In a report concerning the introduction of CAIN on-line services, Jestes (20) provides cost figures for 237 searches run and a composite measure of citation relevancy for 87 searches evaluated. Jestes details all administrative costs including planning and training as well as direct cash-flow costs, and indicates a total average cost per search of $31, and an average cash-flow per search of $14. For the 87 evaluations returned he reports that 3,735 citations (36%) were directly pertinent to the research needs of the user, 2,597 (25%) of the citations were interesting but not directly pertinent, and 3,823 (37%) of the citations were not useful. One hundred and seventy-eight citations were unaccounted for by the respondents.

Recall

Recall assessment determinations present a more formidable obstacle in the assessment of searching performance than precision. Recall however is just as necessary as economic factors, user satisfaction, and relevancy for some evaluations of searching performance.

Some justification for the inclusion of recall is appropriate since its importance for operational on-line systems has been downplayed elsewhere (21,13,22). Cooper acknowledges recall as an important measure in one special case only, which he calls an existence search i.e., one in which a scientist with a new idea searches the literature to find out whether anyone else has thought of the idea first. In this instance he finds recall to be germane but only as auxiliary information to be given to the user for his utility estimate of the search received. If cases such as that described by Cooper

occur (users with a need for existence searches) in institutions providing on-line searches it is appropriate then to consider how well this need is met by the service provided. Furthermore, it is possible to conceive of additional instances in which recall is important beyond that of a scientist with a "new idea." One case is a review article. An individual writing a review publication has a high recall need in approaching a retrieval service. Another case is a research publication. By convention most scientific publications include an introductory section in which the researcher refers to the purpose of his study in context with what has been reported elsewhere. Bibliographic searches initiated to support research publications have a high recall need since, in developing this contextual relationship, the author is open to criticism from reviewers and readers if he ignores (is unaware of) reports of related research with implications relevant to his findings. No existing indexing system uses qualitative search terms such as "good stuff" or "garbage". In their absence, with present systems indicating "aboutness" rather than quality of the document, recall is a legitimate concern for evaluation.

Several studies suggest that a desire for high recall is important to users of information systems. Twenty-six of the 48 searches studied in the AIM - TWX investigation were tagged by the requesters as needing all possible relevant citations (12). In the EARS study 37% of 47 requests evaluated indicated a desire for all possible relevant citations (23). The importance of recall is reinforced by Carmon's research efforts to model the user interface (17). Part of the study concerned a parallel user evaluation conducted at the University of Georgia-Athens, and UCLA. Users of information retrieval services at both institutions ranked lack of a recall measure at the top of their list of least liked characteristics. Another section of the user

evaluation is also seen to concern recall. Twenty-five percent of the users at the University of Georgia and 19% of the users at UCLA indicated that the searches had provided too few citations to be useful. Thirty-five percent of the University of Georgia and 39% of the UCLA users indicated receipt of fewer relevant citations than expected. Lynch and Smith (24) comparing a manual search of the literature with the performance of a profile constructed for a search of the Chemical Titles data base, indicate "To us, as potential users of the service for our everyday research needs . . . accuracy and completeness of retrieval were paramount considerations . . ."

Four studies concerning recall relevant to on-line performance have been reported (12,23,25,16). In both the AIM - TWX and the EARS study Lancaster uses unit cost in minutes of terminal time per relevant citation as one of the evaluation measures. An estimated recall figure is determined for the searches run by biomedical practitioners by having the same search conducted by a trained analyst, with the analyst instructed to achieve as high a recall as possible. No unit cost figures are given for the analyst conducted search. In the AIM - TWX study the majority of the evaluations are from one institution, with a few evaluations from three other participating institutions. The number of analysts conducting recall searches at the institution supplying the majority of evaluations is not indicated. Lancaster's methodology with some modifications appears to offer the most promising avenue for evaluating provision of on-line services. Smetana et al. studied terminal activity and recall in relation to the educational background and experience of 8 searchers investigating the same question with the Lockheed DIALOG system. The variations in search strategies exhibited, and the range of recall levels achieved by the searchers in this study, combined with the variations in search strategy reported here for the CA Condensates insect search reinforce my belief that

recall measures are an integral component of search performance evaluation for some types of questions.

Unit Cost ($) Per Relevant Citation and Recall as Performance Measures

It is suggested that two measures be considered for performance evaluation of on-line searching -- unit cost in dollars per relevant citation and recall. Application of one or both of these measures is dependent upon user expressed needs for SOME relevant citations (unit cost measure) or ALL POSSIBLE relevant citations (unit cost and recall measures).

Unit cost determination as described here differs from that suggested by Lancaster in the AIM - TWX study and Information Retrieval On-Line in its inclusion of a precision measure in relation to off-line prints, as well as a measure of on-line terminal time. Lancaster's suggestion of unit cost expressed as minutes of terminal time used per relevant citation retrieved is based on the supposition that on-line systems will be used for situations in which a few citations are needed in a short time i.e., no off-line printing would be done. In addition he considers that on-line systems will be used for testing strategies for high recall need searches, with the actual search subsequently conducted off-line in a batch mode.

For institutions using on-line systems these suppositions do not appear to be true. It appears that on-line searching is being used to satisfy needs for both kinds of requests (the need for some citations and the need for a high recall search).

Since searchers conducted to satisfy a request for SOME relevant references (i.e. no indicated need for high recall) can include cases in which the requester desires anywhere from one to 50 or more references, an economic factor of terminal time vs. printing costs enters into real use

situations. Actual searching on-line often consists of iterative attempts to develop a best strategy, sometimes including a sampling of citations that would be retrieved by various strategies, and selection of a final strategy or strategies. Where the postings indicate a large number of citations retrieved in this final step (perhaps greater than 30) the searcher may elect to have the results printed off-line. Satisfied with a final strategy the searcher may judge off-line printing at 10¢ per citation printed to be more satisfactory than retrieving every citation on-line for viewing and/or printing in a data base that costs $1.00 per minute for connect time.

It seems then that an efficiency measure for providing a relevant citation must consist not only of the time spent at the terminal but must also include the decision made by the searcher to continue with the iterative process or to conclude the interaction at a certain time and print the results off-line. Such a measure should include a precision assessment of those citations printed off-line since a cost is attached to off-line printing as well as to terminal time. For example the strategy used in Search C of the insect search -- searching only on insect terms with no other limitations -- if applied to an on-line system might use relatively little terminal time. Costs for off-line printing of irrelevant citations would be substantial. A second searcher might use additional terminal time to impose effective restraints on the broad insect term strategy, retrieving approximately the same number of relevant citations off-line but also significantly reducing the printing of irrelevant citations. Lancaster's method for unit cost would show the first search to be more efficient.

Consider instead unit cost per relevant citation expressed in the following manner:

$$C_R = \frac{C_t + C_p}{R}$$

where C_R = Unit cost per relevant citation

C_t = Cost of minutes of terminal time as billed by supplier.*

C_p = Cost of off-line prints

R = Number of relevant citations retrieved

With this expression of unit cost the searcher spending additional minutes at the terminal is not penalized if the costs incurred for the additional minutes result in a comparable or greater reduction in costs associated with off-line printing.

Recall determination is more difficult than unit cost per relevant citation. Recall determinations are also likely to involve a greater economic investment. Yet in those cases in which the searching agency implies to the user that high recall is possible (e.g. by asking varying forms of the question Do you want all possible references? Do you want a broad search?) and the user indicates a high recall need, the searching agency suggests some confidence in its ability to provide high recall results.

Estimated recall values for a sampling of high recall need searches might be determined in the following ways.

A. Institution with one searcher

It is assumed that most strategies involve the intersection of at least

*C_t is determined in this manner to allow comparison of values reported by different search agencies. It excludes Tymshare and other costs which would vary from institution to institution. For individual institutions C_o -- overhead costs (Tymshare, long distance charges to Tymshare node, terminal rental, operation time, etc.) can be added to the numerator to determine actual unit costs for its operation.

two concepts i.e. the occurrence of concept A with concept B, however many synonymous terms are used to express A and/or B (An average of 2.06 ands per question has been reported for 1011 ERIC searches (16), 87% of the profiles studied at the University of Georgia (17) used one or more "ands" per profile). The high recall need search is done in the normal manner. A second search is then conducted on a limited segment of the data base being searched (e.g. three months). The formulation for this search includes only those terms which the searcher has used to express the primary concept of the request (searching on either A or B alone, whichever is considered the primary aspect). If two concepts are equally important in the request the second, time-limited search is done on that concept which the searcher considers was best able to be expressed completely in the real search just completed. For example a requester asks for studies of emotionally disturbed gifted children. "Emotionally disturbed" and "gifted children" might be considered as two equally important concepts. The data base being used may have specific terminology for "gifted children", but terms for the "emotionally disturbed" aspect may be less direct. In this case the second, time-limited search would be done on the term(s) used to express "gifted children."

Each citation from the two searches would be evaluated for relevance by the requester. Unit cost per relevant citation would be determined for the first search in the manner previously described. Estimated recall would be determined by comparing the number of relevant references retrieved in the first search for the time period corresponding to that covered in the second search. Estimated recall would be determined by the formula:

$$R_e = \frac{R_1}{R_2}$$

where R_e = Estimated recall

R_1 = Number of relevant references retrieved in the first search for the time period corresponding to the second search (e.g. Jan.-March 1975)

R_2 = Number of relevant references retrieved in the second search (Jan.-March 1975)

Determination of estimated recall in institutions with one searcher has limitations. Recall in the second search is only as good as the exhaustivity with which the searcher has described that concept selected as primary in the first formulation. For example, in the search on chemical composition of insects, "insects" appears to be the primary concept. A second search for recall on this concept as formulated by Searcher B would be weakened by the lack of exhaustivity in the original formulation, which searched only on "insect" or "insects".

B. Institution with several searchers

With the permission of the requester all on-line searchers would be present for the interview and could participate in the interview process. Alternatively, with the permission of the requester, the interview could be taped and listened to by the other searchers at a later time. The searcher who would normally handle the request would formulate and run the search in the usual manner. Each of the attending searchers would formulate a search strategy for the question and then agree upon the primary concept of the request. A second, time-limited search would be run on the composite of terms suggested by each searcher for the primary concept. Unit cost per relevant citation and estimated recall would be determined in the manner

previously described.

Conclusion

Implementing an evaluation program is not easily accomplished. It involves extra dollars and extra human effort. Performing even the simpler of the two evaluative measures -- unit cost determinations -- is dependent upon the requester's participation in judging citation relevancy. Such participation has been difficult to assure in operations providing free retrieval services. It is likely that it will be even more difficult to elicit user response in operations which charge for searches. Recall evaluations will result in additional on-line and printing costs for duplicate searches. Individuals doing on-line searching may feel threatened if an efficiency measure is attached to their search efforts. Yet, whatever the difficulties attached to such a program, evaluation appears to be an appropriate concern of institutions which deal in information. If the unexamined life is not worth living, it seems that unexamined searching is not worth doing.

TABLE I

Search Characteristics and Costs

	Search A	Search B	Search C
CA Condensates Coverage	vols. 76-81 (1972-74)	vols. 76-81)1972-74)	vols. 78-79)1973)
Number of citations retrieved	544 + 25*	247	1324
Number of minutes at terminal	68	11	N.A.
Cost per minute of terminal time	$ 1.10	$ .95	N.A.
Searching costs	$ 74.80 (6 vols.)	$10.45 (6 vols.)	$ 80.00 (2 vols.)
Print-out Costs	$ 43.52 + 2.00*	$19.76	$ 65.00
Total cost as billed	$120.32	$30.21	$145.00

* 25 citations initially printed, not sent to the user

TABLE 2

Overall Precision and Estimated Recall,
Precision and Recall Adjusted for
Volumes 78-79 CA Condensates,
Unit Costs Per Relevant Citation

	Search A	Search B	Search C
Precision (v. 76-81, 1972-74)	$\frac{101}{544}$ = 18.6%	$\frac{31}{247}$ = 12.6%	$\frac{73}{1324}$ = 5.5%
Estimated Recall (v. 76-81, 1972-74)	$\frac{101}{168}$ = 60.1%	$\frac{31}{168}$ = 18.5%	$\frac{73}{168}$ = 43.5%
Adjusted Precision (vols. 78-79,1973)	$\frac{23}{181}$ = 12.7%	$\frac{3}{64}$ = 4.7%	$\frac{73}{1324}$ = 5.5%
Adjusted Recall (vols. 78-79, 1973)	$\frac{23}{75}$ = 30.7%	$\frac{3}{75}$ = 4.0%	$\frac{73}{75}$ = 97.3%
Unit Cost in minutes of terminal time per relevant citation	.67	.36	N.A.
Unit cost as invoiced costs per relevant citation	$1.19	$ .97	$1.99

FIGURE 1

SEARCH STRATEGY A

I		II		III
Insect Component	AND	Chemical Component	AND NOT	Negated Component
Ant		HORMON#		DNA
Ants		IRIDO#		RNA
Bee		PHEROMON#		ENZYM#
Bees		SEX# and ATTRACTANT#		INSECTICID#
Beetle#		STEROID#		PESTICID#
Bug		TERPEN#		007# (Enzymes)
Bugs		TOXIN#		
Butterfl#		VENOM#		
Cockroach#		COMPO#		
FLY		IDENTIF#		
FLIES		ISOLAT#		
MOTH		PPEPN#		
MOTHS		STRUCTUR#		
SILKWORM#		SYNTH#		
WASP#		032# (Steroids)		
INSECT#		030# (Terpenoids)		
		031# (Alkaloids)		
		033# (Carbohydrates)		
		040# (dyes)		
CA		045# (Fats and waxes)		
PUBLICATION		062# (Essential oils)		
SECTION		021# (General organic chemistry)		
NUMBERS		022# (Physical organic chemistry)		
		023# (Aliphatic compounds)		
		024# (Alicyclic compounds)		
		025# (Non-condensed aromatic compounds)		
		026# (Condensed aromatic compounds)		
		027# (Heterocyclic - one hetero atom)		
		028# (Heterocyclic - more than one hetero atom)		
		029# (Organometallic compounds)		

= truncated search term

FIGURE 2

SEARCH STRATEGY B

I		II
Insect Component	AND	Chemical Component
INSECT		PHEROMONE
INSECTS		LIPIDS
		AMINO and ACID#
		ALLERGEN#
		HORMONE#
		STEROID#
		PROTEIN
		PROTEINS

= truncated search term

Literature Cited

1. CUADRA, CARLOS A. "SDC Experiences With Large Data Bases." Journal of Chemical Information and Computer Sciences 15:1 (Feb.1975) 48-51.

2. HITCHINGHAM, EILEEN E. "Use of Commercially Vended On-Line Data Bases by Academic Libraries." Paper scheduled for presentation at American Society for Information Science, 38th Annual Meeting, October 26-30, Boston, Massachusetts.

3. CUADRA, CARLOS. "Analysis of Man-Machine System Communication in Existing On-Line Retrieval Systems." NSF Contract GN - 41627. Noted in Information, Part 1, 6:4 (April, 1974) 113.

4. PREWITT, BARBARA G. "On-Line Searching Computer Data Bases." Journal of Chemical Documentation, 14:3 (August 1974) 115-117.

5. ELMAN, STANLEY A. "Cost Comparison of Manual and On-Line Computerized Literature Searching." Special Libraries 66:1 (January 1975) 12-18.

6. HOCK, RANDOLPH E. "Providing Access to Externally Available Bibliographic Data Bases in an Academic Library." College and Research Libraries 36:3 (May 1975) 208-215.

7. BENENFELD, A.R., M.E. PENSYL, R.D. MARCUS, J.F. REINTJES. "NASIC at MIT, Phase I Report." Prepared for the New England Board of Higher Education. New England Board of Higher Education, Wellesley, Massachusetts, March 1974. (Appendix A of the NASIC Phase I Report.)

8. LAWRENCE, BARBARA; BEN H. WEIL; MARGARET GRAHAM. "Making On-Line Searching Available in an Industrial Environment." Journal of the American Society for Information Science 25:6 (November-December 1974) 364-369.

9. EMERSON, KATHERINE T. Ed. Proceedings of the Symposium on Measurement of Reference. American Library Association, Library Administration Division, Library Organization and Management Section. ALA, Chicago, Ill., 1974.66 p.

10. SARACEVIC, TEFKO. "The Effect of Question Analysis and Searching Strategy on Performance of Retrieval Systems -- Selected Results from an Experimental Study." In: American Society for Information Science. Annual Meeting. 31st, Columbus, Ohio, October 20-24, 1968. Proceedings. Volume 51: Information Transfer. Greenwood Publishing Corporation, New York, 1968, 185-187.

11. O'DONOHUE, C.H. "Comparison of Service Centers and Document Data Bases - A User's View." Journal of Chemical Documentation, 13:1, (February 1973) 27-29.

12. LANCASTER, F. WILFRED. Evaluation of On-Line Searching in MEDLARS (AIM - TWX) By Biomedical Practitioners. Urbana, Ill.: Illinois University Graduate School of Library Science (Occasional Papers No. 101), February 1972. 21p. (Available as ED-062989).

Literature Cited cont.

13. LANCASTER, F. WILFRED; E.G. FAYEN. Information Retrieval On-Line. Los Angeles: Melville, 1973. 597 p. (A Wiley-Becker & Hayes Series Book)

14. LINEPENSEL, KENNETH C. "On-Line Literature Searching in the Small-to-Medium Sized Research Library. A Comparison of Results in 1973 and 1974 Periods." Presented at a Contributed Paper Session at the 66th Annual Conference, June 8-12, 1975, Special Libraries Association, Chicago, Ill.

15. WANGER, JUDITH. "A View from the Middle: the SDC/ERIC Search Service." ERIC Data Base Users Interchange, Attachment #4, March 1974. Operations Research Inc., Information Systems Division, Bethesda, Md.

16. BOURNE, CHARLES P.; JO ROBINSON; JUDY TODD. Analysis of ERIC On-Line File Searching Procedures and Guidelines for Searching. Institute of Library Research Report #ILR-74-005. Institute of Library Research: Berkeley, California, November, 1974.

17. CARMON, JAMES L. Final Report to the National Science Foundation on Grant GN-41236 to Model the User Interface for a Multidisciplinary Bibliographic Information Network. University of Georgia: Office of Computing Activities, Athens, Georgia (May 1975).

18. RODGER, FRANK B. "Computerized Bibliographic Retrieval Services." Library Trends 23:1 (July 1974) 73-88.

19. AUSTEN, ETHEL.; LAWTON, STEPHEN B. Meeting Ontario's Need for Educational Information: An Evaluation of the SDC/ERIC On-Line Bibliographic Search Service. Ontario Institute for Studies in Education, Toronto, March 1972. (Available as ED 077531)

20. JESTES, EDWARD C. CAIN On-Line Project Report for the Period October-December 1973. Unpublished report, University Library, University of California, Davis. June 1974.

21. DAMMERS, H. F. "The Ecomonics of Computer-Bases Information Systems: A Review." Journal of Documentation, 31:1 (March 1975) 38-45.

22. COOPER, WILLIAM S. "On Selecting a Method of Retrieval Effectiveness. Part I. The Subjective Philosophy of Evaluation." Journal of the American Society for Information Science 24:2 (March-April 1973) 87-100.

23. LANCASTER, F. WILFRED, and R.L. RAPPORT, and J.K. PENRY. "Evaluating the Effectiveness of an On-Line Natural Language Retrieval System." Information Storage and Retrieval 8:5 (1972) 223-245.

24. LYNCH, J.T. and G.D.W. SMITH "Scientific Information by Computer." Nature 230:5290 (March 19, 1971) 153-156.

25. SMETANA, FREDERICK O.; MARY ANN FURNISS; T. ROBERT POTTER. A Study of the Relative Effectiveness and Costs of Computerized Information Retrieval in the Interactive Mode. North Carolina Science and Technology Research Center Report # STR-505. Research Triangle Park, N.C.: North Carolina Science and Technology Research Center (November 1974)